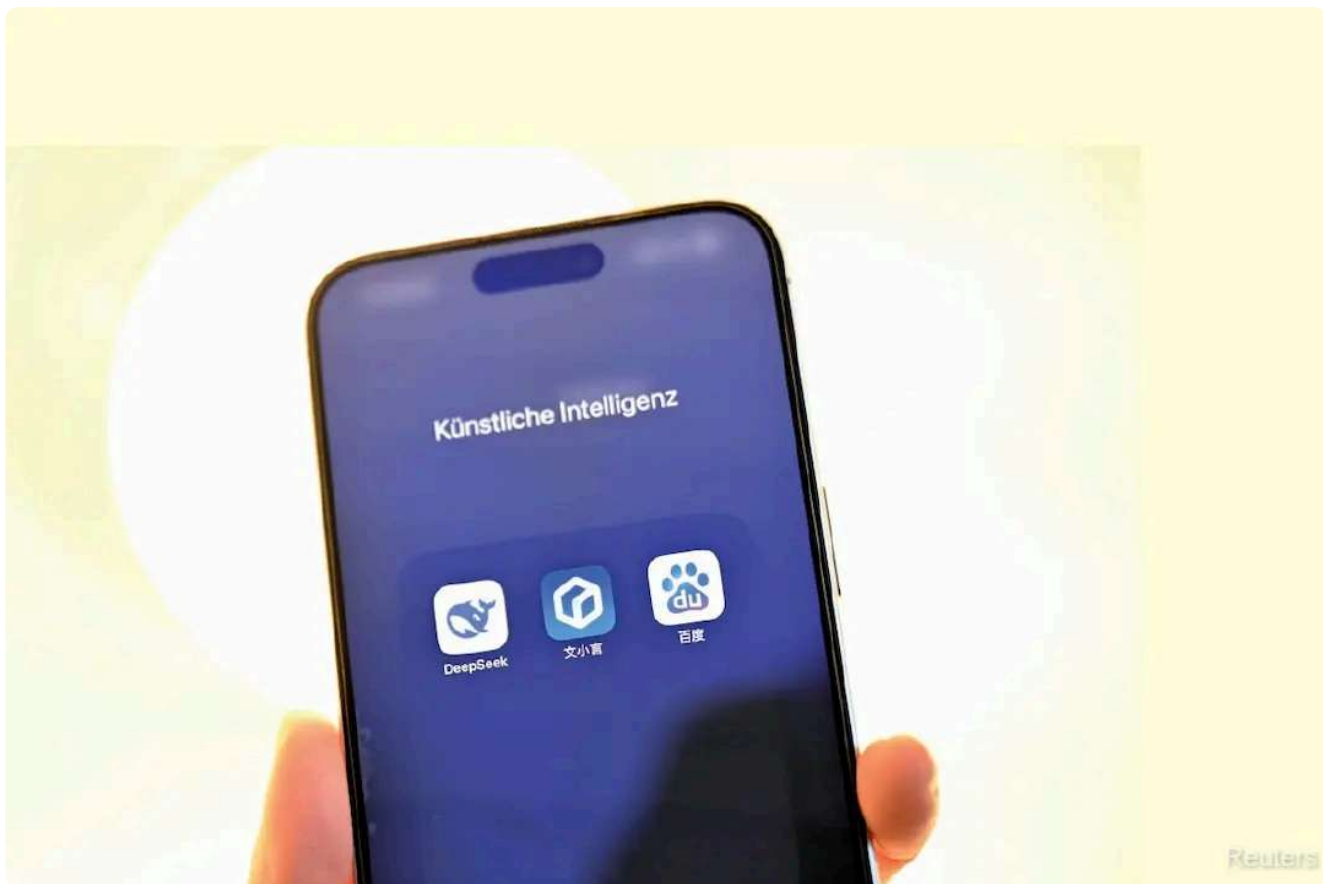


DigitalEdge

Ideas: Is Malaysia ready to seize on China's open-source AI revolution?

By Dr Pieter E Stek / The Edge Malaysia

12 May 2025, 12:00 am



This article first appeared in Digital Edge, The Edge Malaysia Weekly on May 12, 2025 - May 18, 2025

The emergence of the DeepSeek R1 large language model (LLM) in January this year was a watershed for generative artificial intelligence (Gen AI) technology, marking an important milestone from a technological, commercial and political perspective. DeepSeek demonstrated

that advanced LLMs could be trained at just a fraction of the cost previously thought and that this could be done outside of Silicon Valley.

While notable in light of the ongoing US-China technology rivalry and abruptly ending the data centre rally of the Malaysian stock market, the emergence of DeepSeek has far more profound implications for Malaysia's AI landscape.

The first important point to note is that DeepSeek's LLMs are largely based on open-source technology and that models like DeepSeek R1 are either open-source or open-weight, allowing them to be downloaded and modified freely.

Being open source has significant implications for both the technological and commercial trajectory of LLMs.

Chinese technology companies such as Baidu, Alibaba and Tencent have been active in developing open-source AI models for many years. Their strategy, supported by Chinese universities and the government, can be seen as an "open innovation" model aimed at accelerating research and development and leapfrogging past the US.

However, Chinese companies are not the only ones investing in open-source AI. Meta and Google have also released open-weight LLMs for competitive reasons.

A common strategy in technology businesses is to try and "commoditise the complement". If a business is a large user of a product such as Gen AI, rather than using a proprietary model such as ChatGPT, it can be smart to invest in open-source alternatives. Even if the business still uses proprietary LLMs, the availability of a good open-source model erodes the pricing power of a key supplier, such as OpenAI.

A similar strategy was followed by Oracle, which produces servers and networking equipment. Oracle supported the open-source Linux operating system as a way to reduce the pricing power of Microsoft's Windows operating system.

Regardless of the underlying motivations, the fact that high-quality open-weight LLMs are now available means that Malaysia can access them at far lower cost than before.

For the government, this means that it can now run its own LLMs without having to transfer sensitive data to commercial third parties or foreign countries, giving it greater data autonomy.

For companies, open-weight LLMs have levelled the commercial playing field, with start-ups in Malaysia now having access to the same core LLMs as start-ups in China and the US.

Yet, the emergence of Chinese AI has also highlighted a different problem, one that is cultural. Chinese LLMs are known to be trained to repeat the Chinese Communist Party (CCP)'s version of history and its political views, thus conforming to the censorship system in mainland China.

Even if LLMs are not purposely censored, they contain the biases of the texts with which they have been trained. If the texts are primarily in English, then models carry Western cultural viewpoints and biases.

Fortunately, it appears that LLMs can be retrained with relative ease. Just as Chinese LLMs are given guardrails to make them loyal CCP members, another open-source project has already shown that DeepSeek R1 can be post-trained to remove these perceived biases.

For countries like Malaysia, this experience highlights the importance of developing sufficient domestic capacity to localise, train and post-train LLMs for local conditions. Models that do not incorporate Malaysia's racial and religious sensitivities, social hierarchies or slang might not only underperform but also create potentially harmful content.

Some of the capacity to develop LLMs already appears to be present in Malaysia, as in January, local start-up Mesolitica published the open-source MaLLaM LLM, which has a deeper understanding of the subtleties of Bahasa Malaysia than mainstream LLMs like ChatGPT.

However, it is unclear if Malaysian policymakers are fully aware of both the potential of open-source AI and the importance of developing LLMs locally.

Although the National AI Roadmap makes little mention of open source, which is understandable given that it was drafted "long ago" (in 2021), neither do documents from the new National AI Office (NAIO).

While it is difficult to predict the next phase in AI's development, the open-source nature of the current generation of LLMs means that Malaysia has a golden opportunity to catch up with the technology leaders.

To seize this opportunity, Malaysia should update its policies to accommodate the new reality of much smaller and cheaper LLMs. Aside from allowing these models to be adopted more easily, it also makes Gen AI more accessible to small and medium enterprises, as well as for local deployment — for example, in rural areas without reliable internet access.

Malaysia should also expand its capacity to develop LLMs, making them more useful for local languages and more sensitive to local culture. Investments in LLM training could be seen as a

public good and anchored at local universities, thus nurturing local talent and advancing local R&D.

Finally, Malaysia should host its own models to ensure national data autonomy. LLMs can be used to collect large amounts of valuable information which, instead of being used by foreign firms, should be stored and used by local organisations.

Pieter E Stek is a senior lecturer at the Asia School of Business

Save by [subscribing](#) to us for your print and/or digital copy.

P/S: The Edge is also available on [Apple's App Store](#) and [Android's Google Play](#).